Рассчёт правдоподобия гипотезы подгонке родословий в книге Бытие

КОСТЮКОВ Ю. О.*, к. ф.-м. н.

ACM Reference Format:
Костюков Ю. О.. 2025. Рассчёт правдоподобия гипотезы подгонке родословий в книге Бытие. In . ACM, New York, NY, USA, 3 pages.

УСЛОВИЯ И ДОПУЩЕНИЯ

Имеется выборка из $n = 20$ **точно известных** возрастов смерти людей, живших в глубокой древности.

Среди них $K = 15$ возрастов оканчиваются на 0 или 5 (делятся на 5).

Предполагаем, что каждый возраст независим от других (корреляции между возрастами отсутствуют).

Вероятность того, что случайно взятое число делится на 5, равна

$$p = \frac{|\ \{0, 5\}\ |}{|\ \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}\ |} = \frac{2}{10}.$$

Итак, будем проверять две гипотезы.

H_0 **(честная запись).** Приведённые в родословной возрасты точные и реальные, они не были подогнаны тем, кто их записал в текст. Математически это значит, что последняя цифра каждого возраста из родословной равна 0 или 5 с вероятностью $p = \frac{2}{10}$.

H_1 **(подгонка).** Данные в родословной были подогнаны так, чтобы быть «красивыми», т.е. чтобы они (как минимум) делились на 5. Математически: вероятность того, что возраст оканчивается на 0 или 5, является неизвестным параметром p, который не обязательно равен $\frac{2}{10}$.

ПОШАГОВЫЙ РАСЧЁТ С ПОЯСНЕНИЯМИ

Вероятность (p-value) полученного наблюдения крайнего результата при H_0.

Под «крайним» понимаем «не менее 15» успешных исходов.

$$\begin{aligned} P_{H_0}(K \geq 15) &= \sum_{k=15}^{20} \binom{20}{k} p^k (1-p)^{20-k} \\ &= \sum_{k=15}^{20} \binom{20}{k} \left(\frac{2}{10}\right)^k \left(1 - \frac{2}{10}\right)^{20-k} \\ &\approx 1.8 \times 10^{-7} \ = \ 0.000\,018\%. \end{aligned}$$

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

, ,

© 2025 Association for Computing Machinery.

Если записи честные, то шанс получить настолько «круглую» картину — примерно 1 к 5 547 038. Такое событие крайне редкое.

Оценка максимального правдоподобия для H_1.

Гипотеза H_1 говорит: «пусть вероятность того, что возраст кратен 5, *неизвестна*, и мы подберём её, чтобы она делала наблюдаемые 15 успехов наиболее вероятными».

Для биномиального распределения оценка максимального правдоподобия равна

$$\hat{p} = \frac{K}{n} = \frac{15}{20}.$$

Отношение правдоподобий. Это число показывает, во *сколько раз* наши данные «лучше сочетаются» с одной гипотезой, чем с другой.

$$L(H_0) = L(p = 0.2) = \binom{20}{15} p^{15}(1-p)^5 = \binom{20}{15}\left(\frac{2}{10}\right)^{15}\left(1-\frac{2}{10}\right)^5$$

$$L(H_1) = L(\hat{p}) = \binom{20}{15} \hat{p}^{\,15}(1-\hat{p})^5 = \binom{20}{15}\left(\frac{15}{20}\right)^{15}\left(1-\frac{15}{20}\right)^5$$

$$\Lambda = \frac{L(H_0)}{L(H_1)} = \frac{\binom{20}{15}\left(\frac{2}{10}\right)^{15}\left(1-\frac{2}{10}\right)^5}{\binom{20}{15}\left(\frac{15}{20}\right)^{15}\left(1-\frac{15}{20}\right)^5} = \left(\frac{4}{15}\right)^{15}\left(\frac{16}{5}\right)^5 \approx 8.2 \times 10^{-7}$$

Зачем оно нужно? Так как $\Lambda \approx \frac{1}{1\,215\,400}$, данные примерно в 1 215 400 раз правдоподобнее при версии *«возраста подогнаны»*, чем при версии *«возраста честные и* $p = \frac{2}{10}$*»*.

Учёт сложности модели через информационный критерий Акаике.

Критерий Акаике (AIC) равен

$$\mathrm{AIC} = -2\ln L + 2k,$$

где k — число свободных параметров.

$$\mathrm{AIC}(H_0) = -2\ln L(H_0) + 2\cdot 0 \approx 31.22,$$
$$\mathrm{AIC}(H_1) = -2\ln L(H_1) + 2\cdot 1 \approx 5.20,$$
$$\Delta\mathrm{AIC} = \mathrm{AIC}(H_0) - \mathrm{AIC}(H_1) \approx 26.02.$$

Что это значит простыми словами? H_1 «платит» штраф за один дополнительный параметр, но всё равно даёт объяснение данных *значительно* лучше (разница больше 10 — сильное подтверждение).

ИТОГИ

Во-первых, даже при крайне благоприятном предположении увидеть не меньше 15 «круглых» возрастов из 20 — событие с вероятностью около 0.000 018%, что очень мало само по себе.

Во-вторых, отношение правдоподобий показывает, что наблюдаемые данные более чем в миллион раз более правдоподобно согласуются с гипотезой, что *«цифры подогнали»*.

В-третьих, информационный критерий подтверждает, что более разумно предполагать подгонку цифр: хотя это вынуждает нас добавить один свободный параметр (вероятность p), этот параметр окупается, потому что оно радикально повышает объяснительную способность модели.

Вывод. Статистические доводы подавляюще поддерживают гипотезу о том, что указанные 20 возрастов были искусственно «округлены» и не отражают честных демографических данных.

СПИСОК ЛИТЕРАТУРЫ